



Mehr als 20 Jahre Kompetenz und Erfahrung -
lernen Sie uns kennen.

Herzlich willkommen bei saracus!

Hadoop Entwickler Kurs zweitägiges Intensivseminar

Ein Seminar der DWH academy

DWH academy Seminar: Hadoop Entwickler Kurs (2 Tage)

Der 2-tägige Hadoop Entwickler Kurs vermittelt Ihnen die erforderlichen Kenntnisse und Fähigkeiten, um Ihre eigenen Anwendungen für Hadoop entwickeln zu können. Unser Ziel ist, dass Sie so schnell wie möglich produktiv werden.

Hadoop ist eine verteilte Software-Lösung, um große Datenmengen kostengünstig und schnell zu analysieren. Der Kurs vermittelt Ihnen die entscheidenden Grundlagen von Hadoop und dessen MapReduce Modells, dem verteilten Dateisystem HDFS, die Rolle der anderen Hadoop Tools und wie mit diesen Werkzeugen erfolgreich Anwendungen erstellt werden können. Sie werden das Leistungsspektrum von Hadoop kennen und bewerten lernen, um auf diese Weise Einsatzfelder für diese Technologie in Ihrem Unternehmen identifizieren können. Darüber hinaus werden Sie in die Lage versetzt, Ihre eigenen Hadoop-Lösungen zu implementieren.

An wen richtet sich das Seminar?

Das Seminar richtet sich an Entwickler und Architekten, die in der täglichen Praxis mit der Verarbeitung grosser Datenmengen für Analysezwecke konfrontiert sind. Aber auch Entscheider mit Programmierkenntnissen können sich auf diese Weise einen fundierten Eindruck über Hadoop verschaffen und so die Eignung von Hadoop für gegebene Aufgabenstellungen im Big Data-Umfeld konkret bewerten.

Voraussetzungen

Kenntnisse im Bereich der Programmierung (Java, Python oder gleichwertiges), mit der Linux Kommandozeile und Datenbanken/SQL sollten vorhanden sein.

Die Seminarleitung

Das Seminar wird von Ulrich Hebestreit durchgeführt. Herr Hebestreit hat Mathematik studiert und beschäftigt sich seit 30 Jahren mit allen Fragen und Aufgaben zum Thema Software-Engineering. Sein besonderer Schwerpunkt ist seit über 20 Jahren die Architektur, Konzeption und Realisierung von Software-Systemen zur performanten Verarbeitung großer Datenmengen. Seit 1994 ist Herr Hebestreit für die saracus consulting als DWH-Architekt und Chef-Software-Engineer tätig. Dabei nutzt er die Erfahrung aus über 100 Projekten zur Vorbereitung und Durchführung seiner Seminare. Seit zwei Jahren beschäftigt er sich mit dem Thema Hadoop.



Terminplanung

Weitere aktuell geplante Termine der DWH academy finden Sie auf unserer Webseite www.dwh-academy.com.

Teilnahmegebühr

Die Teilnahmegebühr für dieses zweitägige Seminar beträgt 1.750 Euro in Deutschland / 2.100 CHF in der Schweiz zuzüglich der gesetzlichen MwSt. Die Seminarunterlagen, das Teilnahmezertifikat, Mittagessen und Erfrischungen in den Pausen sind in der Teilnahmegebühr enthalten.

Agenda – Tag 1

Ab 8.45 Uhr

Empfang mit Kaffee und Tee

09.00 Uhr

Begrüßung durch die Seminarleitung

09.15 Uhr – 12.00 Uhr

Einführung in Hadoop

- Was ist Hadoop und welche Probleme es lösen kann
- Die Komponenten von Hadoop, deren Rolle und Funktionalität
- Verstehen und Benutzen des HDFS und anderer Dateisysteme
- Das Hadoop Ecosystem
- Ausführen von Jobs auf Hadoop
- Laden von Daten in das Hadoop Cluster

Die MapReduce Programmierung

- Der Datenfluss durch eine MapReduce Anwendung unter Verwendung des klassischen WordCount Algorithmus
- Eine Übersicht der Java MapReduce API und die Anatomie einer Anwendung
- Übung: Java MapReduce Entwicklung mit Eclipse, Testen mit MRUnit und Ausführen von MapReduce Jobs
- Hadoop Streaming zur Programmierung von map und reduce mit Ruby und Python etc.
- Übung: Den WordCount Algorithmus mit Hadoop implementieren

12.00 Uhr – 13.00 Uhr

Mittagessen

Kaffee- und Teepausen finden nach Bedarf statt!

13.00 Uhr – 17:00 Uhr

MapReduce Vertiefung

- Verwendung von Combinern, um den IO Overhead zu reduzieren
- Erstellen von eigenen Key-Value Formaten mit Writables und WritableComparables
- Verwendung von Partitionierern und Comparatorern zur benutzerdefinierten Suche
- Der Secondary Sort Algorithmus
- Übung: Verwendung des Secondary Sort Algorithmus
- Dateiformate: Eigebaute und benutzerdefinierte Formate. Kompressionsmöglichkeiten und das Problem der teilbaren Dateiformate
- Verteilte Dateisysteme: HDFS und andere Dateisysteme
- Zähler und Logdateien
- Wiederverwendbare Mapper mit ChainMapper
- Aufgabenplanung: Die Queue, Fair und Capacity Planer
- Der Verteilte Cache und dessen Verwendung
- Joins: Map-Side und Reduce-Side joins
- Übung: Indexierung des Twitter Verkehrs

ca. 17.00 Uhr

Ende des ersten Seminartages, Zusammenfassung der Ergebnisse

Agenda – Tag 2

Ab 8.45 Uhr

Empfang mit Kaffee und Tee, Ausgabe der Seminarunterlagen

09.00 Uhr

Begrüßung durch die Seminarleitung

09.15 Uhr – 12.00 Uhr

Data Warehousing mit Hive

- Was ist Hive und warum würde ich es nutzen?
- Übung: Ausführen von Hive und grundlegende Abfragen
- Die Abfragesprache Hive (HiveQL: Hive Query Language) an Beispielen
- Übung: Die HiveQL Konzepte an praktischen Übungen verstehen
- Hive und Relationale Datenbanken
- Erweitern von Hive durch benutzerdefinierte Funktionen (UDF: user-defined functions)
- Übung: Integrieren einer UDF in Hive
- Erweitern von Hive durch neue Dateiformate
- Übung: Unterstützen eines benutzerdefinierten Formates (SerDe: record format) in Hive
- Hive unter der Haube: Verstehen wie Hive MapReduce benutzt
- Hinweise zum Aufsetzen des Hive Metadata Repositories
- Tipps und Tricks zu Hive

12.00 Uhr – 13.00 Uhr

Mittagessen

Kaffee- und Teepausen finden nach Bedarf statt!

13.00 Uhr – 17:00 Uhr

Datenfluss Programmierung mit Pig

- Was ist Pig und warum würde ich es nutzen?
- Pig für Datenflüsse und Hive für Abfragen
- Übung: Ausführen von Pig und grundlegende Datenflüsse
- Pig Latein: Die Sprache von Pig anhand von Beispielen
- Übung: Die Pig Konzepte an praktischen Übungen verstehen
- Erweitern von Pig durch benutzerdefinierte Java Funktionen (UDF: user-defined functions)
- Erweitern von Pig durch Python und JavaScript
- Übung: Erweitern von Pig durch UDFs
- Pig unter der Haube: Verstehen wie Pig MapReduce benutzt
- Tipps und Tricks zu Pig
- Übung: Implementierung eines komplexen Datenflussprozesses mit Pig

Job Verwaltung mit Oozie

- Einführung in Oozie
- Erstellen von Arbeitsabläufen mit Oozie
- Übung: Einsatz eines Mahout Recommenders

Zusammenfassung

- Wiederholung des Gelernten
- Wie geht es weiter? Referenzen und Quellen zum Weitermachen

ca. 17.00 Uhr

Diskussion, Zusammenfassung und Ende des Seminars

- Ja**, ich nehme an dem Seminar
„Hadoop Entwickler Kurs“ teil.

Firma: _____

Name: _____ Vorname: _____

Funktion: _____

Straße: _____

PLZ/Ort: _____

Tel.-Nr.: _____ Fax-Nr.: _____

Email: _____

Datum: _____ Unterschrift: _____

**Die Seminargebühr für dieses zweitägige Seminar beträgt 1.750 Euro in
Deutschland / 2.100 CHF in der Schweiz pro Teilnehmer.**

Gebuchter Seminartermin (bitte angeben): _____

Die Liste der Termine finden Sie unter www.dwh-academy.com. Ihre Anmeldung sollte
spätestens 14 Tage vor dem Seminartermin erfolgen.

Gebuchter Seminarort (bitte kreuzen Sie an):

- Baden-Dättwil(CH) Münster (D)
 Ich benötige ein Hotel!

Nach Ablauf der Anmeldefrist erhalten Sie von uns eine Anmeldebestätigung und
eine Rechnung sowie eine Anfahrtsskizze. Bitte überweisen Sie den
Rechnungsbetrag vor Veranstaltungsbeginn.

**Bitte senden Sie diese Seite per Fax an +49 (0251) 98721-26
oder per Post an: saracus consulting GmbH
„Hadoop Entwickler Kurs“
Hafenweg 46, D- 48155 Münster**

Voraussetzung:

Berücksichtigen Sie bitte, dass eine Mindestteilnehmerzahl von 4 Personen vorgesehen ist. Ansonsten kann das Seminar nicht stattfinden.

Anmeldung:

Bitte nehmen Sie Ihre Abmeldung schriftlich vor. Selbstverständlich ist eine Vertretung des angemeldeten Teilnehmers möglich. Alle genannten Preise verstehen sich zzgl. der gesetzl. MwSt.

Stornierung:

Bei Stornierung bis zu 14 Tage vor Veranstaltungsbeginn erheben wir eine Bearbeitungsgebühr von 50% der Gebühr. Bei späteren Absagen wird der gesamte Beitrag fällig.

Seminarabsage:

Sollten wir aus wichtigen Grund (z. B. Erkrankung des/der Referenten) gezwungen sein, den Kurs abzusagen, so teilen wir Ihnen dieses umgehend mit. Wir werden Ihnen in diesem Fall einen Ersatztermin anbieten. Passt dieser nicht zu Ihrer Terminplanung, erhalten Sie die bereits gezahlte Seminargebühr in voller Höhe zurück. Darüber hinausgehende Ansprüche bestehen nicht.



Für Rückfragen steht Ihnen gerne das
DWH-academy-Team der
saracus consulting unter
+49 (0251) 98721-0 zur Verfügung.
Email: info@dwh-academy.com
Internet: www.dwh-academy.com